

The Good Samaritan that wasn't: voluntary monitoring under the (draft) Digital Services Act

Aleksandra Kuczerawy

2021-01-12T14:25:00

On 15 December 2020, the European Commission released the long awaited [proposal for the Digital Services Act](#) (DSA), amending the E-Commerce Directive 2000/31. First impressions (see [here](#) and [here](#)) of the proposal mention that the DSA introduces Good Samaritan protection into the EU intermediary liability regime. A possibility of introducing such a protection was discussed earlier (see Joan Barata [here](#)). Upon further examination, however, it is clear that the new protection is yet something else.

Protecting “Good Samaritans”

In the United States, internet intermediaries – among others, social media platforms like Twitter and Facebook – enjoy “Good Samaritan” protection per [Section 230](#) (47 U.S.C. § 230). Section 230(c) provides that online intermediaries should not be held liable for any voluntary actions taken in good faith against certain types of objectionable content. In particular, it refers to content that “*the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected.*” The aim of this provision was to overrule [Stratton-Oakmont v. Prodigy](#), where an online service provider was found liable for defamatory content uploaded by third parties, because it had tried to detect and remove the objectionable content, but had failed to do so completely. Section 230 also provides that intermediaries shall not be treated as the publisher or speaker of any information provided by another information content provider.

Section 230 essentially protects intermediaries when they act with good intentions to restrict access to or availability of content. But it also protects intermediaries when they do not act against such content, regardless of whether they have knowledge of it or not (provided that this is not content excluded from the scope of Section 230 (c), such as content that violates federal criminal law, e.g. certain child abuse material or terrorist content). The broad protection has been confirmed by extensive case law (see [blog](#) of Eric Goldman). Section 230 is considered as a crucial provision that enabled the development of the internet as we know it today (see Jeff Kosseff, [The Twenty Six Words That Created the Internet](#)).

While Section 230 is subject to increasing criticism in the United States, this will not be discussed further here (for an overview of proposals to amend Section 230 see [here](#)).

No “Good Samaritan” protection under the E-Commerce Directive

In the European Union, the [E-Commerce Directive](#) (ECD) does not explicitly protect internet intermediaries involved in good faith measures against illegal or inappropriate content. Instead, the Directive focuses on knowledge of illegal content, either actual, or constructive.

Per Article 14 ECD, hosting providers can benefit from a liability exemption for hosted speech on condition that they (1) do not have actual knowledge of illegal activity or information and, as regards claims for damages, are not aware of facts or circumstances from which the illegal activity or information is apparent (constructive knowledge); (2) upon obtaining such knowledge or awareness, they act expeditiously to remove or to disable access to the information.

Article 15 ECD prohibits Member States from introducing obligations of general monitoring for intermediaries. However, both specific monitoring obligations and voluntary monitoring on the intermediaries’ own initiative are allowed. But voluntary monitoring could be tricky, as it could lead to awareness of facts or circumstances from which an illegal activity or information is apparent, and therefore to obtaining constructive knowledge (see more [here](#)). As a result, voluntary monitoring could lead to loss of the liability exemption for hosting providers. Nevertheless, the European Commission (EC) has already for quite some time been trying to encourage hosting providers to take up a more proactive approach to content moderation on their services.

Towards Good Samaritan protection in the EU?

The EC has stimulated voluntary removal of content in, for example, the 2016 [Code of Conduct](#) on hate speech, the 2017 [Communication on Tackling Illegal Content](#), or the 2018 [Recommendation](#) on Measures to Effectively Tackle Illegal Content Online.

Inviting private companies to voluntarily monitor, remove or otherwise decrease the visibility of undesirable content raises freedom of expression concerns (see more [here](#) on “interference by proxy”). At the same time, private companies are, strictly speaking, not directly bound by human rights instruments, and the safeguarding of rights or freedoms. An invitation to voluntarily police content also circumvents the general monitoring prohibition of Article 15 ECD, as there is no obligation to monitor, only encouragement. But how to successfully convince intermediaries to engage in voluntary measures, if it could deprive them of the immunity offered by Article 14 ECD?

The EC has entertained the idea of introducing a type of European Good Samaritan protection for some time. The idea was most prominently expressed in the 2017 [Communication on Tackling Illegal Content](#), which referred to proactive measures as ‘so-called “Good Samaritan” actions’. The EC argued that undertaking proactive measures would not automatically lead to a loss of the liability exemption

provided for in Article 14 ECD. The reason being that under Article 14 ECD, to claim immunity, hosts already have to act expeditiously when they obtain knowledge or awareness. As I have written [previously](#), the EC's interpretation offered only half of the Good Samaritan protection. Intermediaries in the EU would not lose the immunity if they took voluntary action resulting in the removal of unlawful or infringing content, but they would not be protected if they did not act (either as a result of a conscious decision or by omission).

The DSA update

How do the EC's efforts for Good Samaritan protection translate into the DSA? The Explanatory Memorandum to the DSA clarifies that the rules on intermediaries build on the ECD. The proposal reproduces the old Articles 12-15 ECD (now Articles 3, 4, 5, and 7) and provides some additional clarifications. For example, Recital 18 states that the liability exemptions should not apply when the provider, instead of providing the services neutrally, by a merely technical and automatic processing of the information provided by the recipient of the service, *"plays an active role of such a kind as to give it knowledge of, or control over, that information."* That codifies earlier interpretations, as well as CJEU rulings in [Google France](#) and [L'Oréal v. eBay](#), that providing 'active' services may lead to the loss of immunity if knowledge could be established. Article 6 DSA is where some might argue a Good Samaritan protection could be found. It aims to eliminate existing disincentives towards voluntary own-investigations undertaken by internet intermediaries:

"Providers of intermediary services shall not be deemed ineligible for the exemptions from liability referred to in Articles 3, 4 and 5 solely because they carry out voluntary own-initiative investigations or other activities aimed at detecting, identifying and removing, or disabling of access to, illegal content, or take the necessary measures to comply with the requirements of Union law, including those set out in this Regulation."

In other words, voluntary measures taken by intermediaries on their own initiative should not be the sole reason for the loss of immunity. Recital 25 of the proposal explains that *"the mere fact that providers undertake such activities does not lead to the unavailability of the exemptions from liability,"* if those activities are undertaken *"in good faith and in a diligent manner."* This applies to activities taken to comply with the requirements of EU law, including those set out in the DSA *"as regards the implementation of their terms and conditions."* The latter is an interesting addition because removals on the basis of terms and conditions are easier and faster for hosts, since they award more discretion to remove content that is unwanted but not necessarily illegal. The recital adds that any such activities should not be taken into account when determining if the service is provided in a neutral manner.

Recital 25 concludes that the rule on determination of neutrality does not imply *"that the provider can necessarily rely thereon."* In other words, taking voluntary actions in good faith neither guarantees nor precludes neutrality. The possibility of losing immunity still exists. Could that be the case when the host undertakes voluntary actions but not particularly successful ones? Would unsuccessful voluntary actions

be considered as not undertaken in a “*diligent manner*”? Could it actually discourage hosts from taking one-time voluntary decisions in particular cases if no coherent framework for ‘diligence’ is in place?

Moreover, as specified in Recital 22, to benefit from the liability exemption, providers of hosting services should act expeditiously to remove or to disable access to content upon obtaining actual knowledge or awareness of its illegal character. Such knowledge or awareness can be obtained through own-initiative investigations or notices submitted by individuals or entities in accordance with the DSA. So, even though voluntary actions by providers would not mean automatic loss of immunity, in specific cases, they might lead to knowledge. For example, if a moderator trained to review for one type of illegality (e.g. incitement to violence) looked at a video, but failed to recognize that it contained another type (e.g. defamation). In such a case, not removing content could still result in liability because the host ‘knew’ or ‘should have known’ about the illegality.

More protection but only for more take-downs?

Article 6 DSA is an interesting addition to the EU intermediary liability regime. The provision clearly attempts to incentivize hosting providers to take more initiative against illegal content, but also content that may violate their self-set terms and conditions, without an immediate risk of losing immunity. This is, of course, welcome from the perspective of hosting providers, as it allows them to moderate content without necessarily being considered as “active”. It also furthers policy-makers’ goal to have intermediaries “do more” content policing without violating the prohibition of general monitoring obligation. At the same time, it is questionable whether facilitating more voluntary removals is actually beneficial from the perspective of users and their right to freedom of expression.

Hosting providers who, until now, may have preferred to remain modest in their voluntary monitoring attempts, because of the risk of being considered active providers with knowledge, need no longer have such qualms. The effect of Article 6 DSA, therefore, could be more content removals. It would be safer for the hosting providers to remain cautious and remove more, rather than less, in avoidance of liability. This outcome would be at odds with one of the goals of the DSA proposal: to ensure more protection for fundamental rights online, in particular the freedom of expression (e.g. in Recitals 3, 22, 41, and 42). It could also deprive the statement in Recital 22, that the “*removal or disabling of access should be undertaken in the observance of the principle of freedom of expression,*” of any deeper meaning. At the same time, as pointed out earlier, conducting the voluntary actions in a less than diligent manner may lead to the loss of immunity, possibly creating an obstacle for more ad hoc voluntary action. Is this contradiction introduced in the text on purpose to balance out the two possible outcomes? And if so, does the new provision actually provide any additional legal certainty? Or will the lack of clarity effectively undercut the Commission’s objectives?

Looking ahead

It will be interesting to see if and how Article 6 DSA evolves during the legislative process. It seems that its main outcome could be more removals by hosting providers on their own initiative, rather than increased protection of the right to freedom of expression. In that sense, the effect would be quite different than the one of Section 230. But several questions remain. Most notably, it is not immediately clear from the text if providers that attempted to take voluntary actions, but were unsuccessful, would still be protected. A case-by-case analysis (e.g. looking at the technology used) would be necessary. Hopefully, the final text of Article 6 DSA will bring more clarity. In any case, it might be time we all stop referring to the new protection as a Good Samaritan clause – the DSA does not use the term – as it has very little in common with its Section 230 counterpart.

The author would like to thank Daphne Keller for her helpful feedback.

